JOURNAL OF MULTIVARIATE ANALYSIS 47, 329–338 (1993)

Simultaneous Estimation of Independent Normal Mean Vectors with Unknown Covariance Matrices

K. KRISHNAMOORTHY

University of Southwestern Louisiana

AND

SANAT K. SARKAR

Temple University

Based on independent samples from several multivariate normal populations, possibly of different dimensions, the problem of simultaneous estimation of the mean vectors is considered assuming that the covariance matrices are unknown. Two loss functions, the sum of usual quadratic losses and the sum of arbitrary quadratic losses, are used. A class of minimax estimators generalizing the James–Stein estimator is obtained. It is shown that these estimators improve the usual set of sample mean vectors uniformly under the sum of quadratic losses. This result is extended to the sum of arbitrary quadratic losses under some restrictions on the covariance matrices.

1. INTRODUCTION

Suppose that there are k independent random vectors $X_1, ..., X_k$ such that X_i: $p_i \times 1$ is from the normal population with unknown mean vector μ_i: $p_i \times 1$ and unknown positive definite covariance matrix Σ_i: $p_i \times p_i$, $N_{p_i}(\mu_i, \Sigma_i)$, $i = 1, ..., k$. Further, suppose that we have independent random matrices $S_1, ..., S_k$, where S_i follows Wishart distribution with parameter matrix Σ_i and degrees of freedom n_i ($\geqslant p_i$), $W_{p_i}(n_i, \Sigma_i)$, $i = 1, ..., k$. Assume that X_i's are independent of S_i's. Let $\hat{\mu} = [\hat{\mu}_1', ..., \hat{\mu}_k']'$ be an estimator of $\mu = [\mu_1', ..., \mu_k']'$. We consider the problem of estimating μ under each of the loss functions

$$L_1(\mu, \hat{\mu}) = \sum_{i=1}^{k} (\hat{\mu}_i - \mu_i)' \Sigma_i^{-1} (\hat{\mu}_i - \mu_i) \tag{1.1}$$

Received February 18, 1992; revised February 9, 1993.

AMS 1980 subject classifications: primary 62F15, secondary 62F10, 62F25.

Key words and phrases: minimax estimator, Wishart, normal, loss function, risk function, eigenvalue.

0047-259X/93 $5.00

and

$$L_2(\mu, \hat{\mu}) = \sum_{i=1}^{k} (\hat{\mu}_i - \mu_i)' Q_i (\hat{\mu}_i - \mu_i) \Bigg/ \left[\operatorname{tr} \sum_{i=1}^{k} \Sigma_i Q_i \right], \tag{1.2}$$

where Q_i's are known positive definite matrices. The setting is the same as that for the problem of estimating the mean vectors of several normal populations based on complete sufficient statistics derived from the samples.

The usual estimator $\mathbf{X} = [X_1', ..., X_k']'$ is minimax as well as the best equivariant for μ under the loss (1.1). James and Stein [11] first showed that, when $k=1$ and $p_1 \geqslant 3$, X_1 is inadmissible for estimating μ_1. They indeed found minimax estimators that improve X_1 significantly under the loss (1.1). Since this monumental work, a substantial number of papers that extend this so-called James–Stein result in different directions have appeared; among them are Alam and Thompson [1], Baranchik [2], Berger [3], Bock [6], Efron and Morris [8, 9], Stein [17], and Strawderman [18]. Recently, the idea of shrinkage has been successfully extended to the problem of estimation of the common mean vector of two normal populations (George [10], Krishnamoorthy [12], and Sarkar [15]) and to the problem of estimation of the coefficient vector in Fisher's linear discriminant function (Sarkar and Krishnamoorthy [16]). However, only limited results are obtained in the simultaneous estimation of several normal means with unknown variances. Efron and Morris [7] considered this problem when $p_1 = \cdots = p_k$ and $\Sigma_1 = \cdots = \Sigma_k$ and derived empirical Bayes estimators dominating $\mathbf{X}$ under the loss (1.1). Rao [14] treated this problem in a regression setting. In the context of several independent linear regression models, he was interested in estimating the regression coefficients simultaneously using the criterion of minimum mean dispersion error matrix. He also obtained empirical Bayes estimators dominating $\mathbf{X}$ under the assumption that $p_1 = \cdots = p_k$ and $\Sigma_i = \sigma^2 V$, for all i, where σ^2 is an unknown scalar and V is a known full-rank matrix. Readers are referred to this article of Rao to see examples where the simultaneous estimation problem of the type considered here arise. Berger and Bock [4] considered a similar estimation problem for the special case when $p_1 = \cdots = p_k = 1$ under the loss (1.2). Their proposed shrinkage estimators dominate $\mathbf{X}$ if $k \geqslant 3$ and n_i's are large. When $k=1$, Berger *et al.* [5] showed that the usual estimator X_1 is inadmissible under the loss (1.2).

The problem considered in this paper is more general in form than the problems treated in some of aforementioned papers. We not only allow the dimension to vary from population to population, but also the covariance matrices are assumed to be unknown and different. The present problem can also be regarded as that of estimating the mean vector with a patterned covariance matrix, an extension of Berger and Bock's [4] problem.

The estimators considered here are of the form

$$\hat{\mu}_\phi = \left[1 - \phi(\mathbf{F})\left(\sum_{i=1}^{k} F_i\right)^{-1}\right]\mathbf{X}, \tag{1.3}$$

where $\phi(\mathbf{F})$ is a function of $\mathbf{F} = (F_1, ..., F_k)'$ with $F_i = X_i' S_i^{-1} X_i$, $i = 1, ..., k$. In Section 2, we first prove that $\hat{\mu}_\phi$ dominates $\mathbf{X}$ and hence is minimax, with respect to the loss (1.1), if $p = \sum_{i=1}^{k} p_i > 2$ and ϕ satisfies some conditions. This is a natural multi-population generalization of Baranchik's [2] main result; of course, a new inequality (Lemma 2.2) needs to be developed to establish this generalization. This general result leads to the interesting conclusion that even in the case of an unequal number of observations from several independent multivariate normal populations (possibly of different dimensions) with unknown covariance matrices the usual set of mean vectors can be dominated by its suitable shrikage version.

Finally, we make an attempt to extend our result to the loss (1.2). We indeed show that, if $p > 2$, some estimators from the class (1.3) dominate $\mathbf{X}$ also under the loss (1.2) and a condition on $\mathrm{tr}(\sum_{i=1}^{k} Q_i \Sigma_i)$. We also explain in Remark 2.5 that, when $k = 1$, a condition on the parameter space is in fact necessary for $\hat{\mu}_\phi$ to dominate $\mathbf{X}$ under the loss (1.2).

It should be pointed out that since the distributions of the sample covariance matrices are sensitive to outliers, the dominance of the proposed estimators over $\mathbf{X}$ is questionable when outliers are present. We, however, do not make any attempt here to investigate the robustness aspect of the present results in order to avoid deviating from the main theme of this paper.

2. Main Results

First, we give the following lemmas, which are needed to prove the main results of this paper stated in Theorems 2.1 and 2.2. In the following, $E(\cdot, \cdot)$ denotes the joint expectation of the arguments and $E(\cdot \mid Y)$ denotes the conditional expectation given Y.

Lemma 2.1. *For any positive valued random variable x, we have that $E[x^t\phi(x)]/E(x^t)$ is nondecreasing (nonincreasing) in t if $\phi(x)$ is a nondecreasing (nonincreasing) function of x and the expectations exist.*

Proof. Let $f(x)$ be the density of x. Then,

$$E[x^t\phi(x)]/E(x^t) = \int \phi(x)\, x^t f(x)\, dx / E(x^t). \tag{2.1}$$

Now, the family of densities $\{x^t f(x)/E(x^t)\}$ has monotone likelihood ratio in x. Hence the lemma follows from Lehmann [13, Lemma 2, p. 85].

LEMMA 2.2. *Let $x_1, ..., x_k$ be mutually independent positive random variables. Then, for some nonnegative constants $g_1, ..., g_k$ satisfying $\Sigma_{i=1}^k g_i = 1$, we have*

$$\frac{E[(\sum_{i=1}^k g_i x_i)^t \phi(X)]}{E[(\sum_{i=1}^k g_i x_i)^{t'} \phi(X)]} \geqslant \min_{1 \leqslant i \leqslant k} \left\{\frac{E(x_i^t)}{E(x_i^{t'})}\right\} \tag{2.2}$$

when $t' \leqslant 0$, $t' \leqslant t \leqslant t'+1$ and $\phi(X)$ is a nonnegative, componentwise nondecreasing function of $X = (x_1, ..., x_k)'$, and the indicated expectations exist.

Proof. Without any loss of generality, we can assume that all g_i's are positive. Under the condition on t' and t, since $x^{t-t'}$ is a concave function of x, we have from Jensen's inequality that

$$\left(\sum_{i=1}^k g_i x_i\right)^{t-t'} \geqslant \sum_{i=1}^k g_i x_i^{t-t'}. \tag{2.3}$$

Using (2.3), we note that the left-hand side of (2.2) is

$$\geqslant \frac{E[(\sum_{i=1}^k g_i x_i^{t-t'})(\sum_{i=1}^k g_i x_i)^{t'} \phi(X)]}{E[(\sum_{i=1}^k g_i x_i)^{t'} \phi(X)]} = \sum_{i=1}^k g_i \frac{E[x_i^t \psi_i(x_i)]}{E[x_i^{t'} \psi_i(x_i)]}, \tag{2.4}$$

where

$$\psi_i(x_i) = E\left[\left(1 + \sum_{k \neq i} \frac{g_k x_k}{g_i x_i}\right)^{t'} \phi(X) \mid x_i\right], \tag{2.5}$$

$i = 1, ..., k$. Observe that, for each i, $\psi_i(x_i)$ is nondecreasing in x_i if $t' \leqslant 0$ and $\phi(X)$ is nondecreasing in x_i. Application of Lemma 2.1 then gives that (2.4) is

$$\geqslant \sum_{i=1}^k g_i \frac{E(x_i^t)}{E(x_i^{t'})} \geqslant \min_{1 \leqslant i \leqslant k} \frac{E(x_i^t)}{E(x_i^{t'})}$$

under the conditions on ϕ, t, and t' stated in the lemma. Thus, the lemma is proved.

THEOREM 2.1. *Let $p \geqslant 3$ and $c^* = [\max_{1 \leqslant i \leqslant k}(n_i - p_i + 3)]^{-1}$. Then, the estimator $\hat{\mu}_\phi$ given in (1.3) dominates $\mathbf{X}$ under the loss (1.1) and hence it is minimax if $0 < \phi(\mathbf{F}) \leqslant 2(p-2)c^*$ and $\phi(\mathbf{F})$ is differentiable and nondecreasing in F_i, $i = 1, ..., k$.*

Proof. The risk of $\hat{\mu}_\phi$ under the loss (1.1) is given by

$$\begin{aligned}
R_1(\mu, \hat{\mu}_\phi) &= E\left\{\sum_{i=1}^{k} (\hat{\mu}_i - \mu_i)' \Sigma_i^{-1} (\hat{\mu}_i - \mu_i)\right\} \\
&= E\left\{\sum_{i=1}^{k} (X_i - \mu_i)' \Sigma_i^{-1} (X_i - \mu_i)\right\} \\
&\quad + E\left\{\phi^2 \sum_{i=1}^{k} X_i' \Sigma_i^{-1} X_i \left(\sum_{j=1}^{k} F_j\right)^{-2}\right\} \\
&\quad - 2E\left\{\phi \sum_{i=1}^{k} X_i' \Sigma_i^{-1} (X_i - \mu_i) \left(\sum_{j=1}^{k} F_j\right)^{-1}\right\}.
\end{aligned} \tag{2.6}$$

To show that $\hat{\mu}_\phi$ dominates $\mathbf{X}$, it is sufficient to verify that the difference $R_1(\mu, \mathbf{X}) - R_1(\mu, \hat{\mu}_\phi)$ is $\geqslant 0$, that is,

$$2E \sum_{i=1}^{k} \left\{\phi X_i' \Sigma_i^{-1} (X_i - \mu_i) \left(\sum_{j=1}^{k} F_j\right)^{-1} - \phi^2 X_i' \Sigma_i^{-1} X_i \left(\sum_{j=1}^{k} F_j\right)^{-2}\right\} \geqslant 0 \tag{2.7}$$

for all μ_i and Σ_i. Let $Y_i = \Sigma_i^{-1/2} X_i$ and $V_i = \Sigma_i^{-1/2} S_i \Sigma_i^{-1/2}$, where $\Sigma_i^{-1/2}$ is a symmetric square root of Σ_i^{-1}, $i = 1, ..., k$. As $X_i \sim N_{p_i}(\mu_i, \Sigma_i)$ independently of $S_i \sim W_{p_i}(n_i, \Sigma_i)$, $Y_i \sim N_{p_i}(\theta_i, I_i)$ independently of $V_i \sim W_{p_i}(n_i, I_i)$, where I_i denotes the identity matrix of order $p_i \times p_i$ and $\theta_i = \Sigma_i^{-1/2} \mu_i$, $i = 1, ..., k$. Also, note that $F_i = X_i' S_i^{-1} X_i = Y_i' V_i^{-1} Y_i$ for all i. In terms of these new variables, (2.7) can be expressed as

$$2E \sum_{i=1}^{k} \left\{\phi Y_i' (Y_i - \theta_i) \left(\sum_{j=1}^{k} F_j\right)^{-1} - \phi^2 Y_i' Y_i \left(\sum_{j=1}^{k} F_j\right)^{-2}\right\} \geqslant 0. \tag{2.8}$$

We here observe that, for fixed i, the components y_{ij} of Y_i are independent normal random variables with means θ_{ij} and variance unity. Now, applying Stein's [17] identity, that $Eh(y_{ij})(y_{ij} - \theta_{ij}) = Eh'(y_{ij})$ for any differentiable function h satisfying $E|h'(y_{ij})| < \infty$, to the first term in the left-hand side of (2.8) and under the conditions on ϕ, it can be checked that (2.8) is reduced to

$$\begin{aligned}
E\Bigg\{2(p-2)\,\phi \left(\sum_{i=1}^{k} F_i\right)^{-1} &- \phi^2 \left(\sum_{i=1}^{k} Y_i' Y_i\right)\left(\sum_{i=1}^{k} F_i\right)^{-2} \\
&+ 4 \sum_{i=1}^{k} (\partial\phi/\partial F_i)\, F_i \left(\sum_{i=1}^{k} F_i\right)^{-1}\Bigg\} \geqslant 0.
\end{aligned} \tag{2.9}$$

Let $\mathbf{Y} = [Y_1', ..., Y_k']'$ so that $\|\mathbf{Y}\|^2 = \Sigma_{i=1}^{k} Y_i' Y_i$. Further, let $U_i = Y_i' Y_i / Y_i' V_i^{-1} Y_i$ and $w_i = Y_i' Y_i / \|\mathbf{Y}\|^2$, $i = 1, ..., k$. It is well known that U_i follows chi-squared distribution with $n_i - p_i + 1$ degrees of freedom independently of Y_i, $i = 1, ..., k$. Also, note that $F_i = \|\mathbf{Y}\|^2 w_i U_i^{-1}$, $i = 1, ..., k$. Thus, in terms of these new variables and under the conditions on ϕ, (2.9) holds if

$$2(p-2)\, E\left\{\|Y\|^{-2} E\left[\phi\left(\sum_{i=1}^{k} w_i U_i^{-1}\right)^{-1} - a\phi\left(\sum_{i=1}^{k} w_i U_i^{-1}\right)^{-2}\right] \middle| \mathbf{Y}\right\} \geqslant 0, \tag{2.10}$$

for any a such that $\phi/(2(p-2)) \leqslant a \leqslant c^*$. As $w_i > 0$ (with probability 1 under the distribution of Y_i), $\Sigma_{i=1}^{k} w_i = 1$, and ϕ in nondecreasing in U_i^{-1}, $i = 1, ..., k$, applying Lemma 2.2 (with $t = -1$ and $t' = -2$) to the first conditional expectation in (2.10), we see that (2.10) holds if

$$2(p-2)(c^* - a)\, E\left[\phi\, \|\mathbf{Y}\| \left(\sum_{i=1}^{k} w_i U_i^{-1}\right)\right]^{-2} \geqslant 0. \tag{2.11}$$

As ϕ is bounded above, the proof will be completed if we show that $E[\|\mathbf{Y}\| (\sum_{i=1}^{k} w_i U_i^{-1})]^{-2}$ is finite. Using Jensen's inequality, we get

$$\left(\sum_{i=1}^{k} w_i U_i^{-1}\right)^{-2} \leqslant \left(\sum_{i=1}^{k} w_i U_i\right)^2 \leqslant \sum_{i=1}^{k} w_i U_i^2. \tag{2.12}$$

So, it follows from (2.12) that

$$\begin{aligned} E\left\{E\left[\|\mathbf{Y}\|^{-2}\left(\sum_{i=1}^{k} w_i U_i^{-1}\right)^{-2}\right] \middle| \mathbf{Y}\right\} &\leqslant E\left[\|\mathbf{Y}\|^{-2} \sum_{i=1}^{k} w_i (EU_i^2)\right] \\ &\leqslant E\,\|\mathbf{Y}\|^{-2}\,[\max_{1\leqslant i\leqslant k} EU_i^2]. \end{aligned} \tag{2.13}$$

As U_i^2's independently follow $\chi^2_{n_i - p_i + 1}$, $i = 1, ..., k$, and $\|\mathbf{Y}\|^2$ follows a chi-squared distribution with $p + 2Z$ degrees of freedom, where Z is a Poisson random variable with mean $\sum_{i=1}^{k} (\mu_i' \Sigma_i^{-1} \mu_i)/2$, the right-hand side of (2.13) exists provided $p - 2 > 0$.

Remark 2.1. Clearly, the most obvious choice of ϕ satisfying the conditions of Theorem 2.1 is a constant function. Another choice is $\phi(\mathbf{F}) = c_0/[1 + c_0(\sum F_i)^{-1}]$, with the corresponding shrinkage estimator equal to $\hat{\mu}_\phi = [\sum_{i=1}^{k} F_i/(\sum_{i=1}^{k} F_i + c_0)]\mathbf{X}$, where $0 < c_0 < 2(p-2)\,c^*$. This is a generalization of Example 2 of [2] and the estimators in [1].

Remark 2.2. When we have independent observations $X_{i1}, ..., X_{iN_i}$ from the ith population, define $n_i = N_i - 1$, $N_{(k)} = \max_{1 \leqslant i \leqslant k}\{N_i\}$, $\bar{X}_i = \sum_{j=1}^{N_i} X_{ij}/N_i$, $S_i = \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$, and $F_i = \bar{X}_i' S_i^{-1} \bar{X}_i$, $i = 1, ..., k$.

Let, $\hat{\mu}_\phi = [\hat{\mu}'_{\phi 1}, ..., \hat{\mu}'_{\phi k}]'$ with $\hat{\mu}_{\phi i} = \{1 - \phi N_i (\sum_{i=1}^k F_i)^{-1}\} \bar{X}_i$, $i = 1, ..., k$. Then $\hat{\mu}_\phi$, when $0 < \phi \leqslant 2(p-2) c^*/N_{(k)}$ and ϕ satisfies the other condition of Theorem 2.1, dominates the sample mean vector $\bar{\mathbf{X}} = [\bar{X}'_1, ..., \bar{X}_k]'$ under the loss (1.1). An optimal choice of ϕ for practical purposes is $\phi \equiv (p-2) c^*/N_{(k)}$.

Remark 2.3. The positive part version of $\hat{\mu}_\phi$, that is,

$$\hat{\mu}_\phi = \left(1 - \phi\left(\sum_{i=1}^k F_i\right)^{-1}\right)^+ \mathbf{X},$$

where $a^+ = \max(0, a)$, will improve on $\hat{\mu}_\phi$.

We next consider the estimation of μ under the arbitrary quadratic loss (1.2). For this loss, when $p_1 = \cdots = p_k = 1$, Berger and Bock [4] proposed the estimator given componentwise by $\delta_i(\mathbf{X}, W) = (1 - r(\mathbf{X}, W)/[\|\mathbf{X}\|_W^2 q_i W_i]) X_i$, where $\|\mathbf{X}\|_W^2 = \sum_{i=1}^k [X_i^2/(q_i W_i^2)]$ and $W_i = S_i/(n_i - 2)$. They proved that this estimator, when $k \geqslant 3$, r satisfies some conditions, and the sample sizes are sufficiently large, dominates the usual minimax estimator. When $k = 1$, Berger *et al.* [5] demonstrated that the estimator $\hat{\mu}(X_1, S_1, Q_1) = (1 - c\alpha Q_1^{-1} S_1^{-1}/(X'_1 S_1^{-1} X_1)) X_1$, where α is the minimum eigenvalue of $Q_1 S_1/(n_1 - p_1 - 1)$, dominates X_1 with respect to the loss (1.2) for some appropriate values of c. In the following theorem we show that the estimator $\hat{\mu}_\phi$ dominates $\mathbf{X}$ also under the quadratic loss (1.2) and certain conditions on ϕ and the covariance matrices.

The following lemma is also needed to prove Theorem 2.2.

LEMMA 2.3. *Let $X \sim N_p(\theta, I)$ independently of $S \sim W_p(n, I)$. Also, let $\eta(B)$ denote the maximum eigenvalue of a positive semidefinite matrix B. Then, for a positive real-valued function f,*

$$0 < E(X'BS^{-1}Xf(X'S^{-1}X)) \leqslant \eta(B)\, E(X'S^{-1}Xf(X'S^{-1}X)),$$

provided the indicated expectations exist.

Proof. Choose an orhtogonal matrix R such that $RX = \|X\| e_1$, where e_1 is the column vector whose first component is unity and others are zeroes. Let $B_x = RBR'$ and $A = RSR'$. Also, write A as

$$\begin{pmatrix} a_{11} & A'_{21} \\ A_{21} & A_{22} \end{pmatrix}$$

such that $A_{22}: (p-1) \times (p-1)$ and define $a_{11.2} = a_{11} - A'_{21} A_{22}^{-1} A_{21}$. It is easy to see that

$$\begin{aligned} X'BS^{-1}Xf(X'S^{-1}X) &= \|X\|^2 e'_1 B_x A^{-1} e_1 f(\|X\|^2 e'_1 A^{-1} e_1) \\ &= \|X\|^2 a_{11.2}^{-1} e'_1 B_x (1, -A_{22}^{-1} A_{21})' f(\|X\|^2 a_{11.2}^{-1}). \end{aligned}$$

We now note that $A =^d S$ independently of X, from which we see that $(a_{11\cdot 2}, X)$ is independent of (A_{21}, A_{22}), and $A_{21} \mid A_{22} \sim N_{(p-1)}(0, A_{22})$. Using these results and taking expectation on both sides of the above equation, we then get

$$0 < E(X'BS^{-1}Xf(X'S^{-1}X)) = E(\|X\|^2 a_{11\cdot 2}^{-1} e_1' B_x e_1 f(\|X\|^2 a_{11\cdot 2}^{-1}))$$
$$\leqslant \eta(B)\, E(\|X\|^2 a_{11\cdot 2}^{-1} f(\|X\|^2 a_{11\cdot 2}^{-1})).$$

To obtain the second inequality we used the result that $e_1' B_x e_1 \leqslant \eta(B_x) = \eta(B)$. Thus, we complete the proof by noting that $\|X\|^2 a_{11\cdot 2}^{-1} =^d X'S^{-1}X$.

THEOREM 2.2. *Let λ_i denote the maximum eigenvalue of $\Sigma_i Q_i$, $i = 1, ..., k$, $\lambda = \max_{1 \leqslant i \leqslant k}\{\lambda_i\}$, $p \geqslant 3$, and $c^* = [\max_{1 \leqslant i \leqslant k}\{n_i - p_i + 3\}]^{-1}$. Further, assume that $\operatorname{tr}(\sum_{i=1}^k Q_i \Sigma_i) \geqslant 3\lambda$. Then, the estimator $\hat{\mu}_\phi$ dominates $\mathbf{X}$ under the loss (1.2), if $0 < \phi(\mathbf{F}) \leqslant 2c^*$ and is differentiable and nondecreasing in F_i, $i = 1, ..., k$.*

Proof. Proceeding as in the proof of Theorem 2.1 and after using Stein's [17] identity we can see, under the loss (1.2), that the risk difference $R_2(\mu, \mathbf{X}) - R_2(\mu, \hat{\mu}_c) \geqslant 0$ if and only if

$$\sum_{i=1}^k E\left\{2\phi \operatorname{tr} \Omega_i \left(\sum_{j=1}^k F_j\right)^{-1} - 4\phi Y_i' \Omega_i V_i^{-1} Y_i \left(\sum_{j=1}^k F_j\right)^{-2} \right.$$
$$+ 4(\partial\phi/\partial F_i)\, Y_i' \Omega_i V_i^{-1} Y_i \left(\sum_{j=1}^k F_j\right)^{-1}$$
$$\left. - \phi^2 Y_i' \Omega_i Y_i \left(\sum_{j=1}^k F_j\right)^{-2}\right\} \geqslant 0, \tag{2.14}$$

where $Y_i \sim N_{p_i}(\theta_i, I_i)$ independently of $V_i \sim W_{p_i}(n_i, I_i)$, $\theta_i = \Sigma_i^{-1/2}\mu_i$, and $\Omega_i = \Sigma_i^{1/2} Q_i \Sigma_i^{1/2}$, $i = 1, ..., k$. Now, for each fixed i, applying Lemma 2.3 to the second term in the left-hand side of (2.14) we can see, under the conditions of the theorem, that (2.14) holds if

$$E\left\{2\lambda\phi\left(\sum_{j=1}^k F_j\right)^{-1} - a\phi \sum_{i=1}^k Y_i' \Omega_i Y_i \left(\sum_{j=1}^k F_j\right)^{-2}\right\} \geqslant 0 \tag{2.15}$$

for any a such that $0 \leqslant \phi \leqslant a \leqslant 2c^*$. Define U_i, ω_i, and $\mathbf{Y}$ as in Theorem 2.1. Using the relation that $Y_i' \Omega_i Y_i / Y_i' Y_i \leqslant \lambda_i$, $i = 1, ..., k$, we see that (2.15) is true if

$$\lambda E\left\{2\phi\left(\sum_{i=1}^k w_i U_i^{-1}\right)^{-1} \|\mathbf{Y}\|^{-2} - a\phi\left(\sum_{i=1}^k w_i U_i^{-1}\right)^{-2} \|\mathbf{Y}\|^{-2}\right\} \geqslant 0. \tag{2.16}$$

Thus, applying Lemma 2.2 (with $t' = -2$, $t = -1$) to (2.16) and then proceeding along the same lines as those in the proof of Theorem 2.1, we can complete the proof.

Remark 2.4. In stead of $\hat{\mu}_\phi$ one can also consider a more general class of estimators of the form

$$\hat{\mu}_{\phi c} = \left[1 - \phi\left(\sum_{i=1}^{k} c_i F_i\right)^{-1}\right]\mathbf{X},$$

where c_i's are known positive constants. Theorems 2.1 and 2.2 will still hold with $\hat{\mu}_\phi$ and c^* respectively replaced by $\hat{\mu}_{\phi c}$ and $c^{**} = [\max_{1 \leqslant i \leqslant k}\{c_i(n_i - p_i + 3)\}]^{-1}$.

Remark 2.5. The estimator $\hat{\mu}_\phi$ is independent of the positive definite matrices Q_i's given in the loss (1.2), whereas the estimators suggested by Berger and Bock [4] and Berger *et al.* [5] are dependent on Q_i's. Although these estimators are functions of Q_i's (thus depending on the loss function), they uniformly dominate the usual estimator. As $\hat{\mu}_\phi$ does not improve **X** uniformly, one may be interested in knowing whether there is any estimator that is free of Q_i's and better than the usual estimator. For $k = 1$, the answer is no and it can be proved as follows: Let $\hat{\mu}_1 = (\hat{\mu}_{11}, ..., \hat{\mu}_{1p_1})'$ be an estimator of $\mu_1 = (\mu_{11}, ..., \mu_{1p_1})'$ independent of Q_1. Assume that it dominates $X_1 = (x_{11}, ..., x_{1p_1})'$ under the loss (1.2). This implies that, for some μ_1 and Σ_1,

$$E(X_1 - \mu_1)' Q_1(X_1 - \mu_1) - E(\hat{\mu}_1 - \mu_1) Q_1(\hat{\mu}_1 - \mu_1) = \operatorname{tr}(Q_1 P) > 0,$$

where $P = E(X_1 - \mu_1)(X_1 - \mu_1)' - E(\hat{\mu}_1 - \mu_1)(\hat{\mu}_1 - \mu_1)'$. It is not difficult to verify that $\operatorname{tr}(Q_1 P) > 0$ for any arbitrary positive definite matrix Q_1 if and only if P is a positive semidefinite matrix with at least one positive eigenvalue. This means that $E(x_{1i} - \mu_{1i})^2 > E(\hat{\mu}_{1i} - \mu_{1i})^2$ for at least one i, $i = 1, ..., p_1$, which contradicts the fact that x_{1i} is admissible as an estimator of μ_{1i}, $i = 1, ..., p_1$, under the squared error loss function. Thus we prove that, when $k = 1$, X_1 is admissible in the class of estimators that are independent of Q_1. However, when $k \geqslant 3$, it is plaussible in the class of estimators that are indpendent of Q_1. However, when $k \geqslant 3$, it is plausible that the usual estimator is inadmissible even in this subclass. We are currently investigating this problem and plan to report it separately.

Acknowledgment

The authors gratefully acknowledge some useful comments made by the referees.

REFERENCES

[1] ALAM, K., AND THOMPSON, J. R. (1964). *Estimation of the Mean of Multivariate Normal Distribution*. Technical Report. Indiana University.

[2] BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **41** 642–645.

[3] BERGER, J. (1976). Admissible minimax estimation of a multivariate normal means with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226.

[4] BERGER J., AND BOCK, M. E. (1976). Combining independent normal mean estimation problems with unknown variances. *Ann. Statist.* **4** 642–648.

[5] BERGER, J., BOCK, M. E., BROWN, L. D., CASELLA, G., AND GLESER, K. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5** 763–771.

[6] BOCK, M. E. (1975). Minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **3** 209–218.

[7] EFRON, B., AND MORRIS, C. (1972). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika* **59** 335–347.

[8] EFRON, B., AND MORRIS, C. (1973). Stein's estimation rule and its competitors—An empirical Bayes approach. *J. Amer. Statist. Assoc.* **70** 311–319.

[9] EFRON, B., AND MORRIS, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4** 11–21.

[10] GEORGE, E. I. (1991). Shrinkage domination in a multivariate common mean problem. *Ann. Statist.* **19** 952–960.

[11] JAMES, W., AND STEIN, C. (1960). Estimation with quadratic loss. In *Proceedings, Fourth Berkeley Symp. Math. Statist. Probab.*, Vol. 1, pp. 361–379. Univ. of California Press, Berkeley.

[12] KRISHNAMOORTHY, K. (1991). On a shrinkage estimator of a normal common mean vector. *J. Multivariate Anal.* **40** 109–114.

[13] LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*. Wiley, New York.

[14] RAO, C. R. (1975). Simmultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics* **31** 545–554.

[15] SARKAR, S. K. (1992). *Shrinkage Domination of Some Usual Estimators of the Common Mean of Several Multivariate Normal Populations*. Technical Report No. 91-16, Temple University, Philadelphia.

[16] SARKAR, S. K., AND KRISHNAMOORTHY, K. (1991). *Estimation of a Function of Multivariate Normal parameters*. Technical Report No. 91–1, Department of Statistics, Temple University.

[17] STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

[18] STRAWDERMAN, W. E. (1973). Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. *Ann. Statist.* **1** 1189–1194.